Earth Sciences Department

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

What

Environmental forecasting

Why

Our strength …
- … research …
- … operations …
- … services …
- … high resolution …

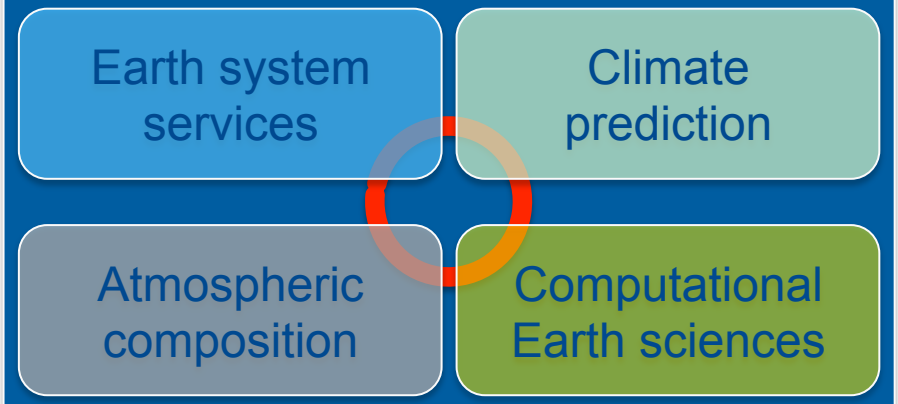
How

Develop a capability to model air quality processes from urban to global and the impacts on weather, health and ecosystems

Implement climate prediction system for subseasonal-to-decadal climate prediction

Develop user-oriented services that favour both technology transfer and adaptation

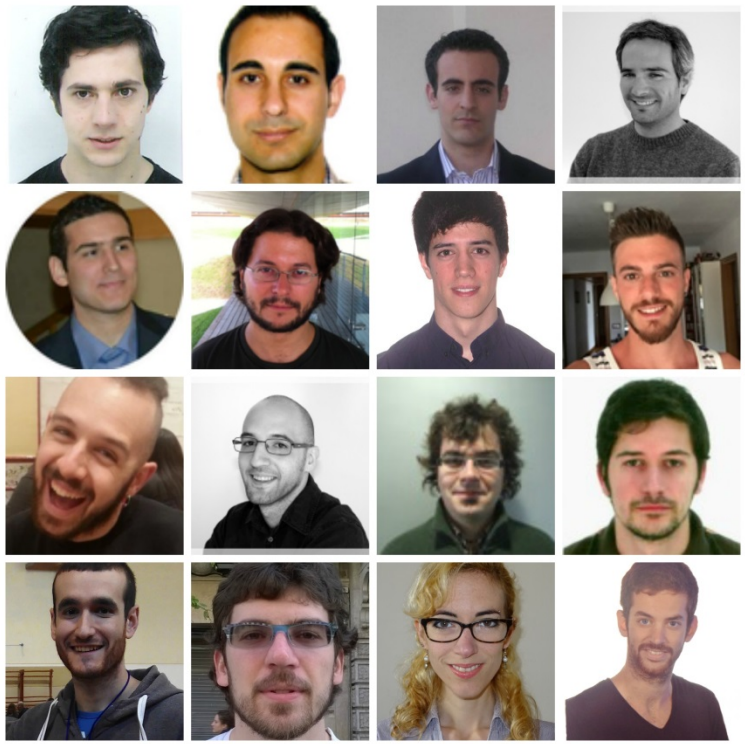
Use cutting-edge HPC and Big Data technologies for the efficiency and user-friendliness of Earth system models



Members

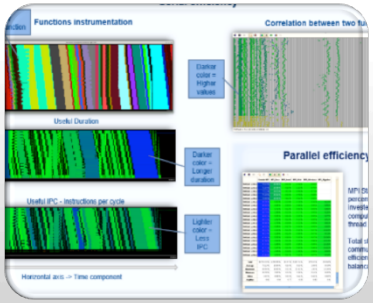
- Multidisciplinary team with different IT profiles

- Currently, 17 members
  - 2 Managers
  - 10 engineers
  - 2 Postdoc
  - 1 PhD student
  - 1 Master student
  - 1 Intern



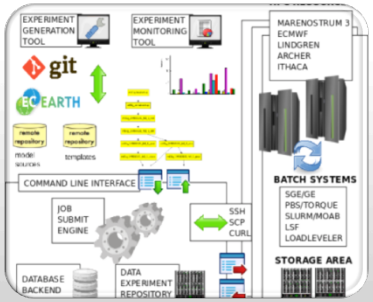
- Foster internal BSC interdepartmental collaboration, specially Computer Science to apply their research

Structure



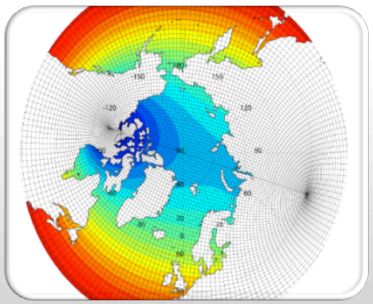
Performance Team

- Provide HPC Services
- Apply new computational methods



Models and Workflows Team

- Development of HPC user-friendly software framework
- Support the development of climate and atmospheric research software



Data and Diagnostics Team

- Big Data in Earth Sciences
- Provision of internal and external data services
- Visualization

Funding and projects

- Mainly competitive founds (FP7, H2020)
  - Projects from Climate or Atmospheric Sciences
    - SPECS, PRIMAVERA, RESILIENCE, CALIOPE…
  - Computational Earth Sciences projects
    - IS-ENES2
    - ESiWACE
- Collaborations with other research institutes or institutions
  - ECMWF, IPSL, CERFACS
  - AEMET
  - NCEP

BSC Collaboration

- Internal collaboration with Computer Science department
  - Performance tools
  - Programming Models
  - Workflows and Distributed Computing
  - Storage systems

- Applying CS research to Earth System models
  - Porting ESM's to OmpSs
  - Deploying ESM's in novel architectures
  - Porting workflows to new tools

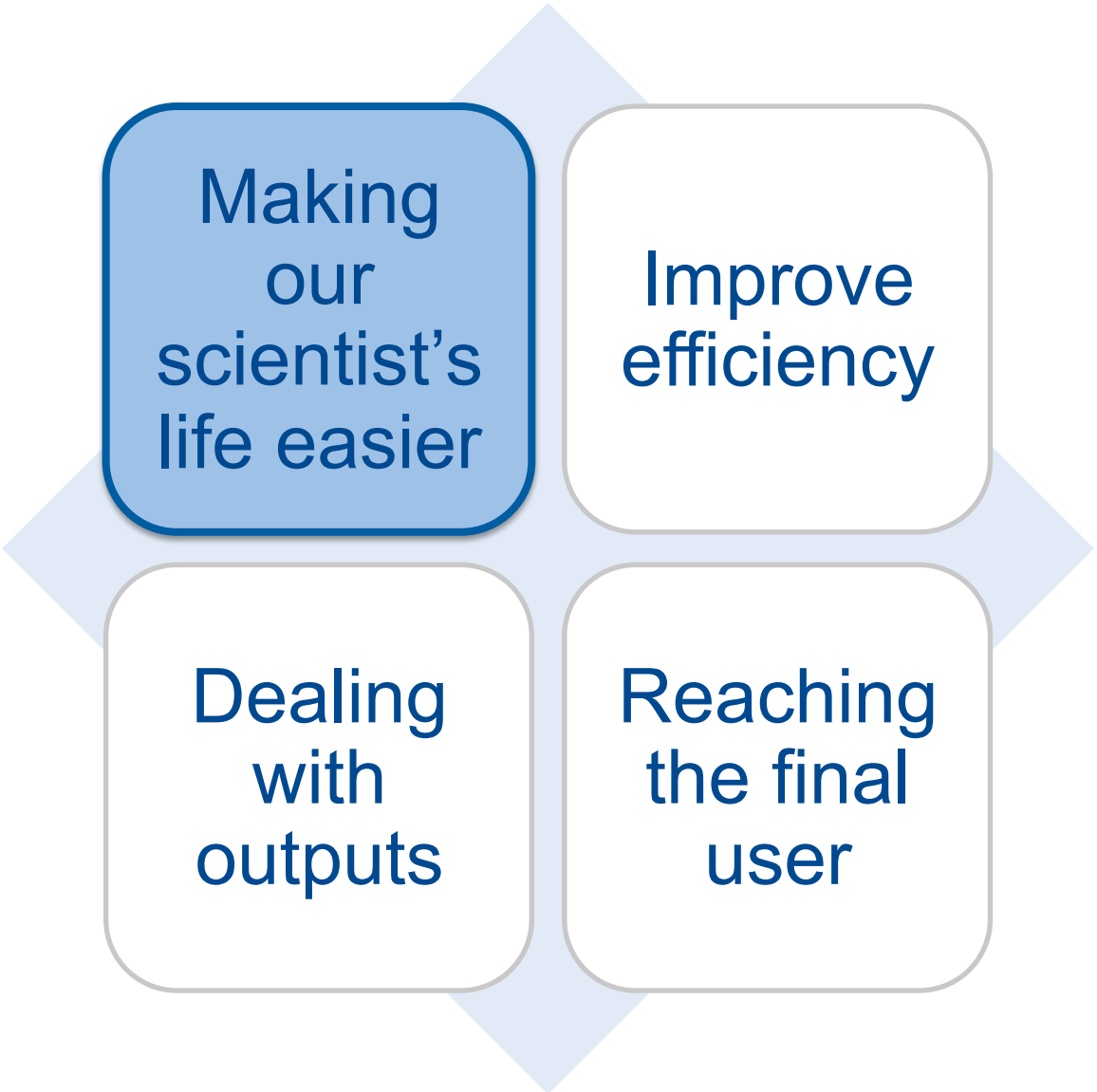
www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

OUR CHALLENGES

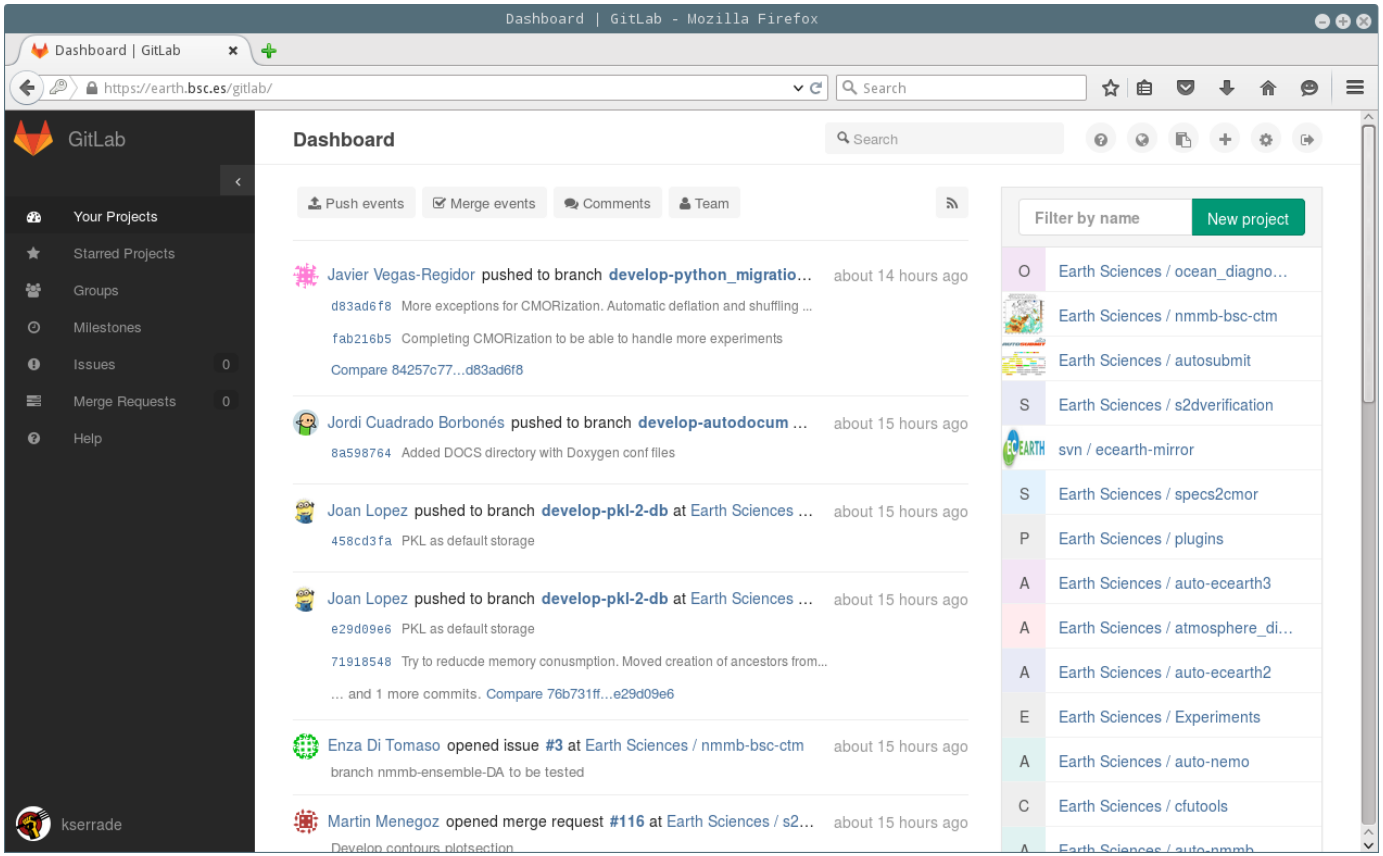
Some challenges



Making our scientist's life easier

- Earth System Models are NOT easy to use
  - Sometimes hard to develop
  - Sometimes hard to deploy
  - Sometimes hard to run
  - Sometimes hard to work with produced outputs

Gitlab + Unit Testing

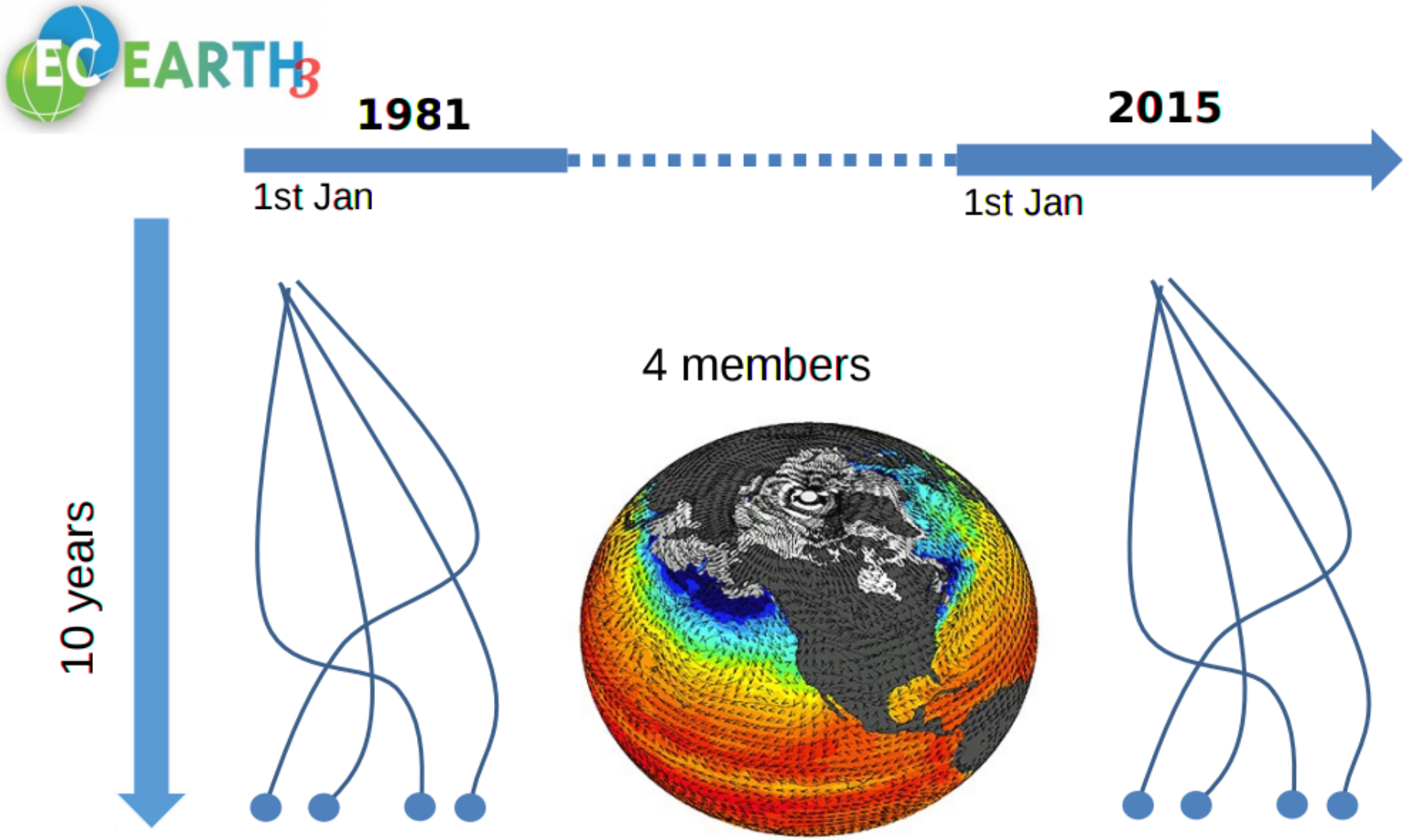


- Deploying a gitlab server
- Developing a unit testing strategy (for some projects)
- Continuous training scientists to use these tools

Easybuild for Earth Sciences

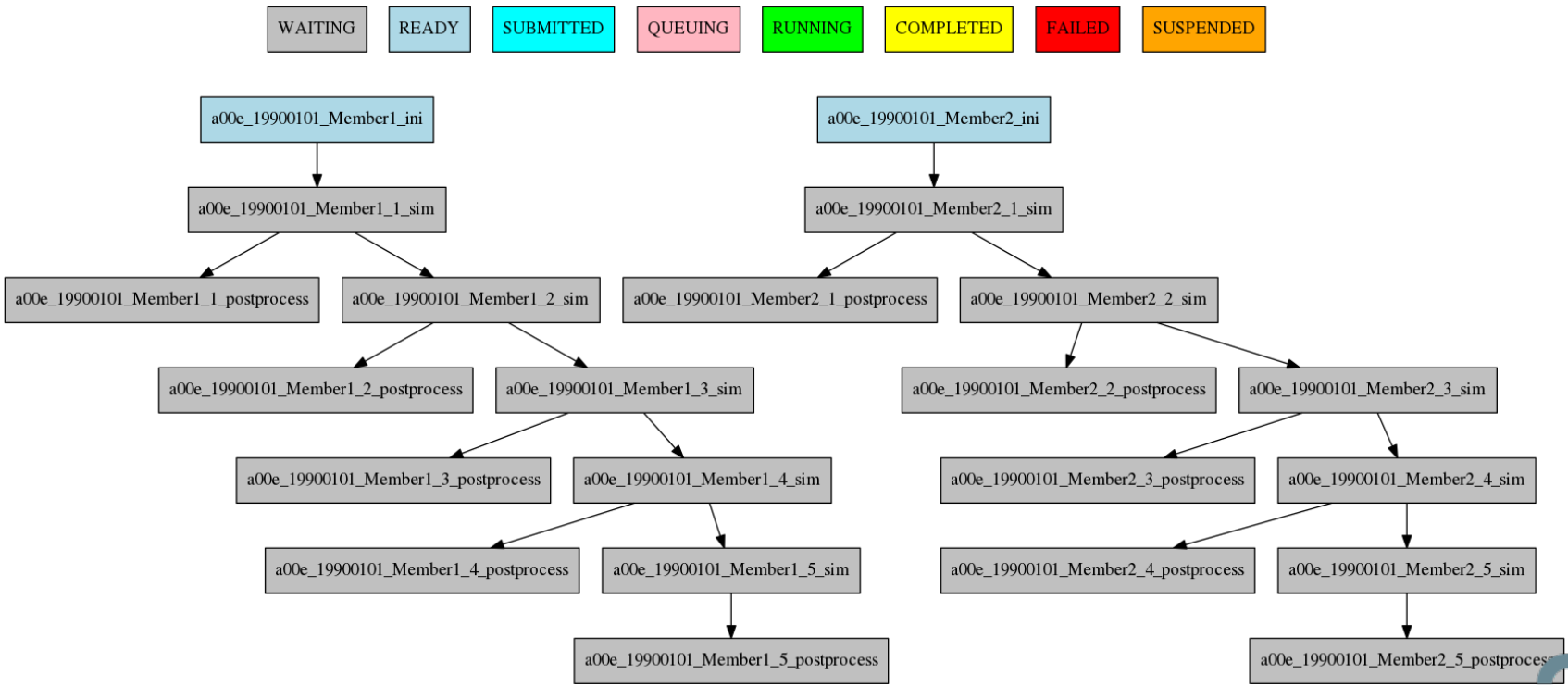
- EasyBuild is a software build and installation framework that allows you to manage (scientific) software on High Performance Computing (HPC) systems in an efficient way
  - https://hpcugent.github.io/easybuild/
- Developed for libraries and general purpose software
  - Handles dependencies and modules
- We ported to our models and libraries (work in progress)
  - NEMO/3.6-r6664-foss-2015a-BSC
  - XIOS/r858-foss-2015a
  - NMMB-BSC/2.0.4-foss-2015a
  - ESMF/7.0.0-foss-2015a

Running a decadal experiment



Autosubmit

- Autosubmit
  - A versatile tool to manage Weather and Climate Experiments in diverse Supercomputing Environments
  - https://pypi.python.org/pypi/autosubmit



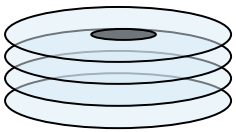
Autosubmit

- Automatization: Preparing and running, postprocessing and output transfer, all managed by Autosubmit. No user intervention needed.
- Provenance: Assigns unique identifiers to each experiment and stores information about model version, configuration options, etc
- Failure tolerance: Automatic retrials and ability to repeat tasks in case of corrupted or missing data.
- Versatility: Currently runs EC-Earth, NEMO and NMMB/BSC models on several platforms

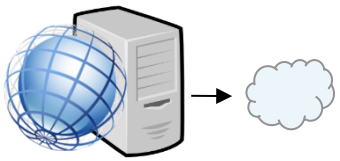
s2dverification

- s2dverification
  - Set of tools to verify forecasts through the computation of typical prediction scores against one or more observational datasets or reanalyses

s2dverification package



LOCAL STORAGE



ESGF NODE
or
OPeNDAP SERVER

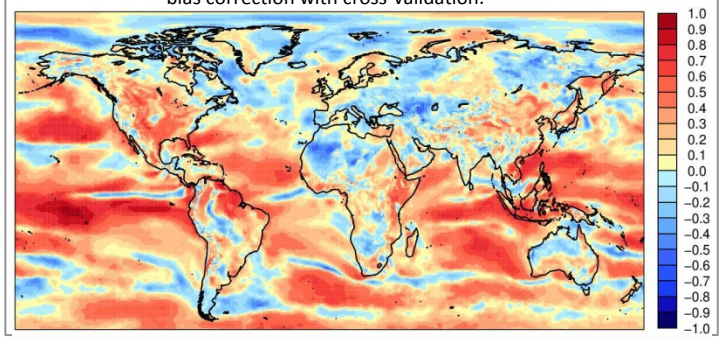
- Supports datasets stored locally or in ESGF (OPeNDAP) servers.
- Exploits multi-core capabilities
- API available
- Collects observational and experimental datasets stored in multiple conventions:
  - NetCDF3, NetCDF4
  - Supports specific folder and file naming conventions.

BASIC STATISTICS

SCORES
EXTREMES

PLOTS

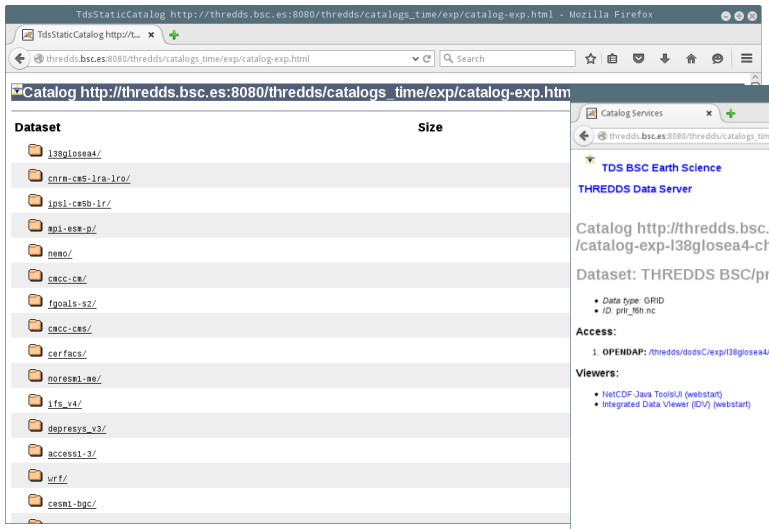
Anomaly Correlation Coefficient. 10M Wind Speed ECMWF S4 1 month lead with start dates once a year on first of November and Era-Interim in DJF from 1981 to 2011. Simple bias correction with cross-validation.

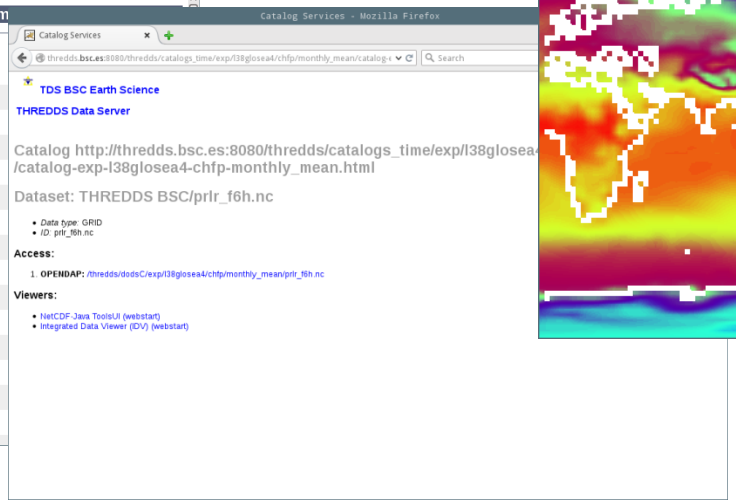


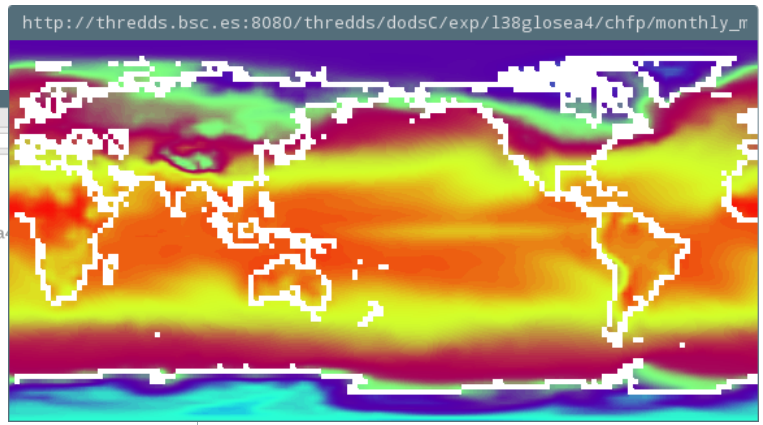
Data management

- Our users don’t really know where the data is
  - With internal THREDDS server we can mix storages
  - We can use other features

- Drawback
  - Data has to be stored with a precise format (data and metadata)
  - A technician working exclusively with data

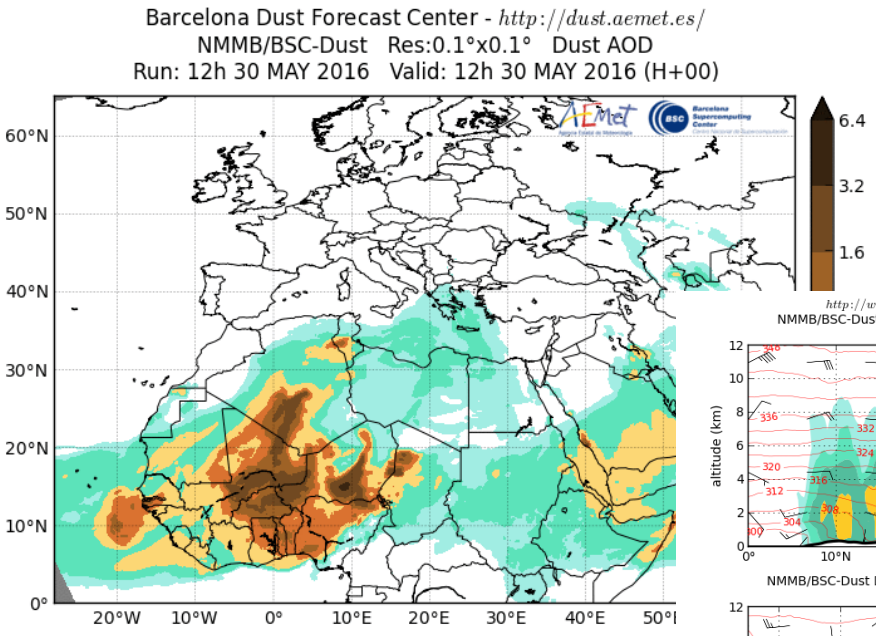


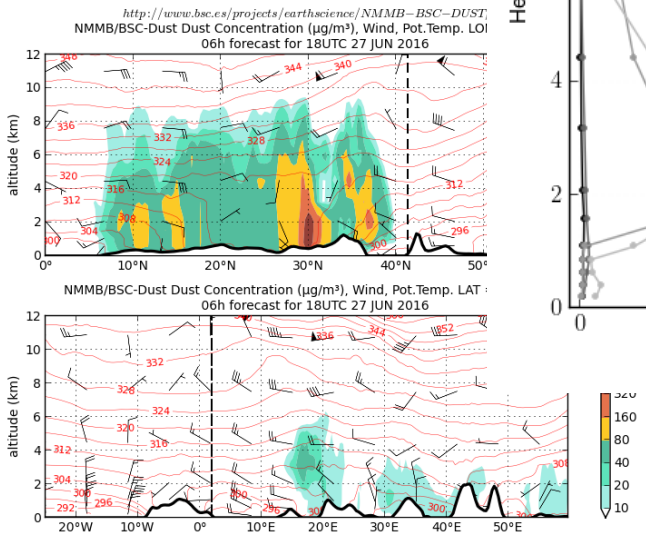


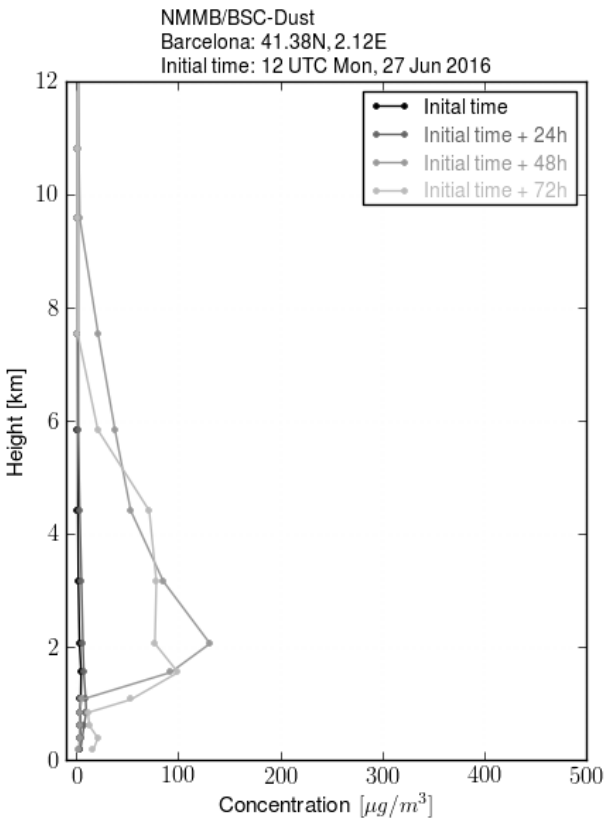


Visualization

- A python package, MapGenerator for atmospheric composition visualization
  - Python and GrADS

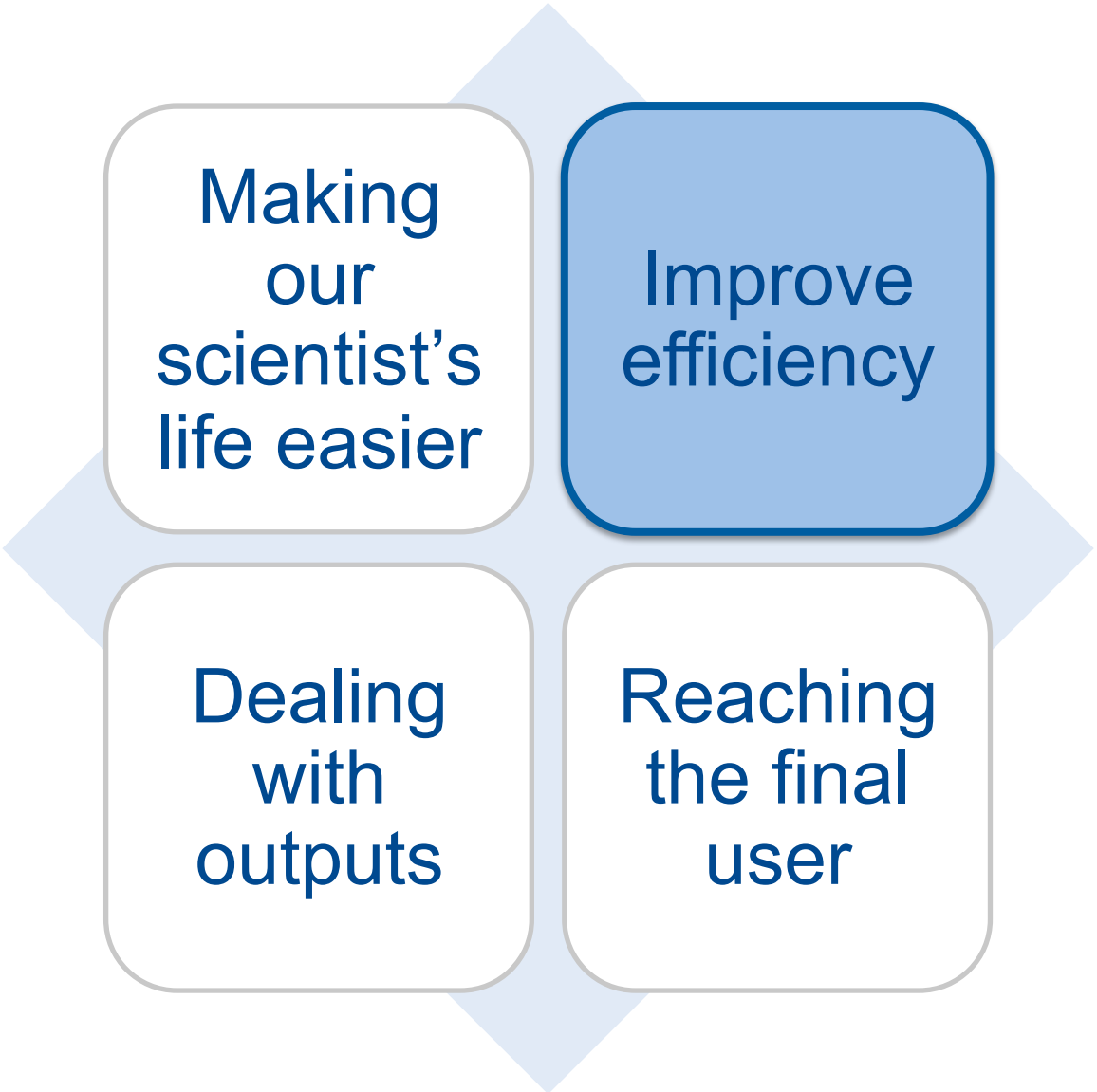






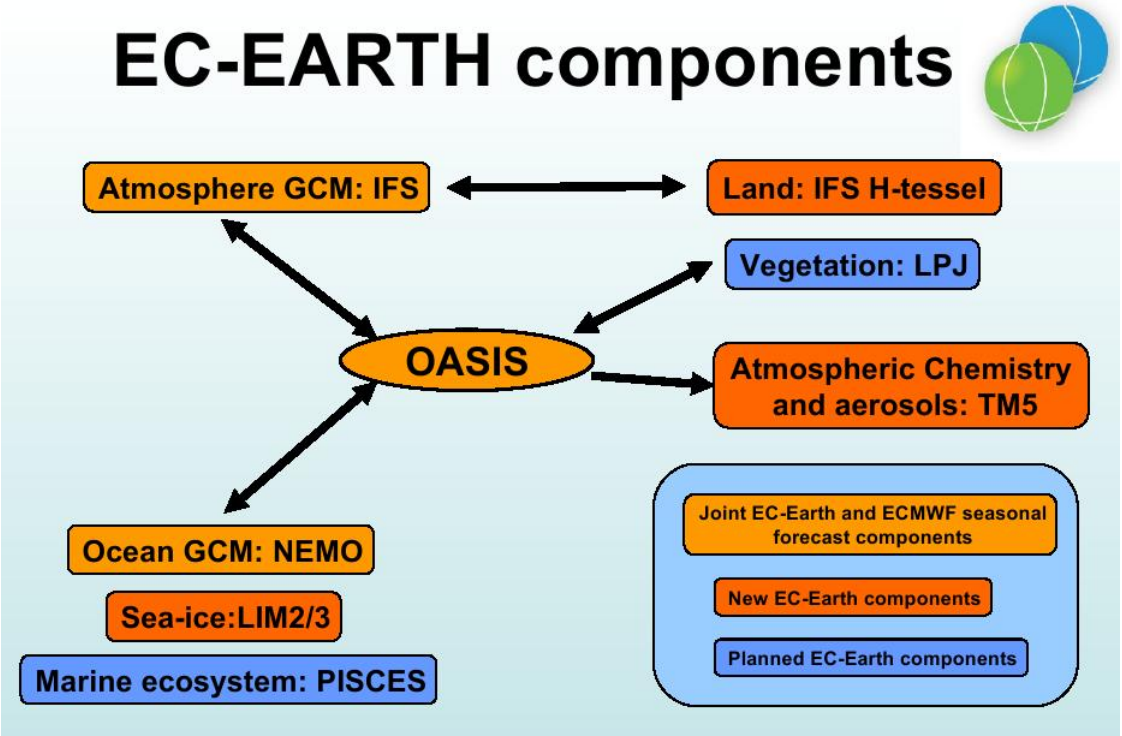
F., and E. Terradellas (2012): MapGenerator: a toolbox to process and visualize air quality datasets, EuroSciPy 2012, Belgium, Brussels.

Some challenges

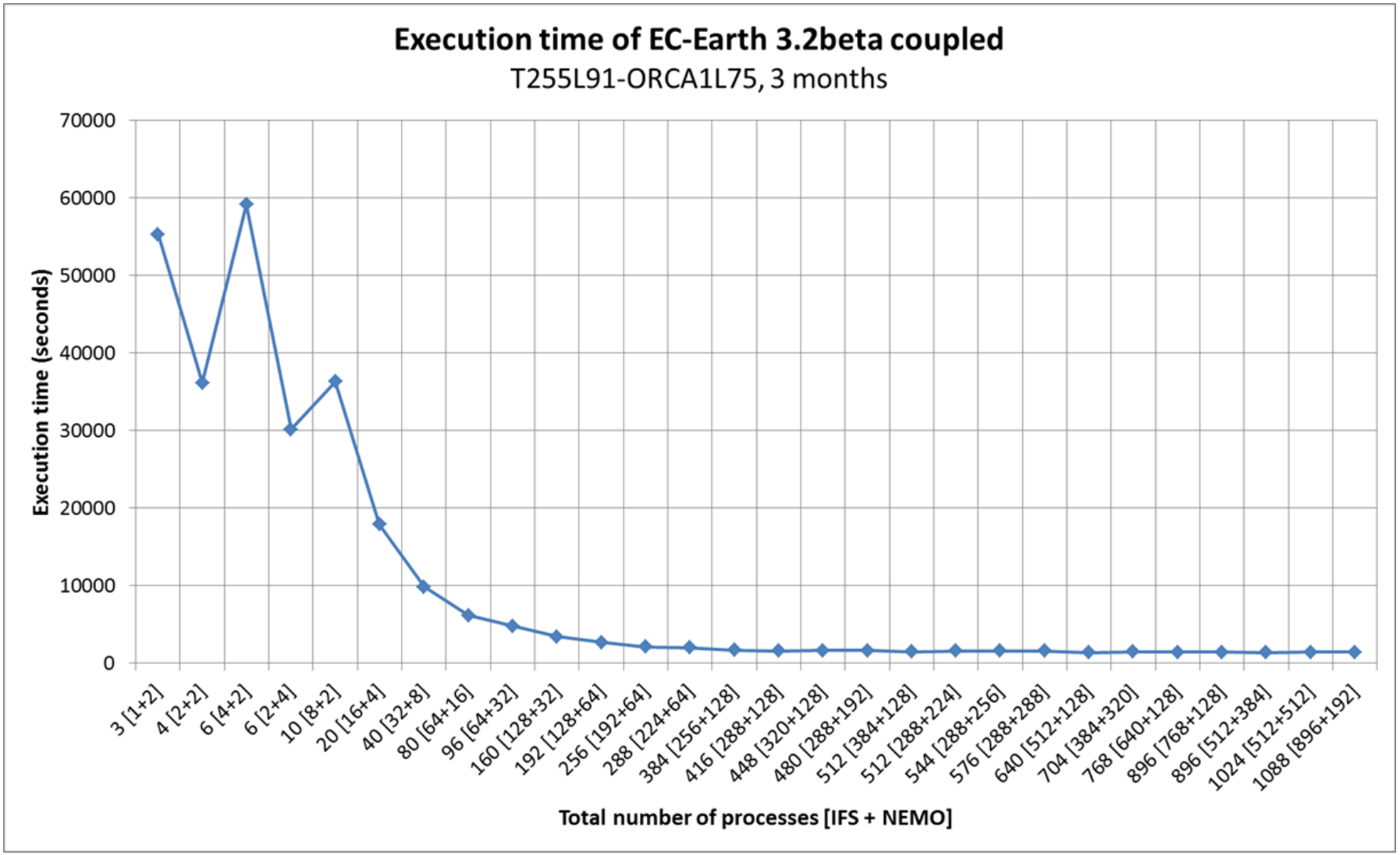


Performance analysis

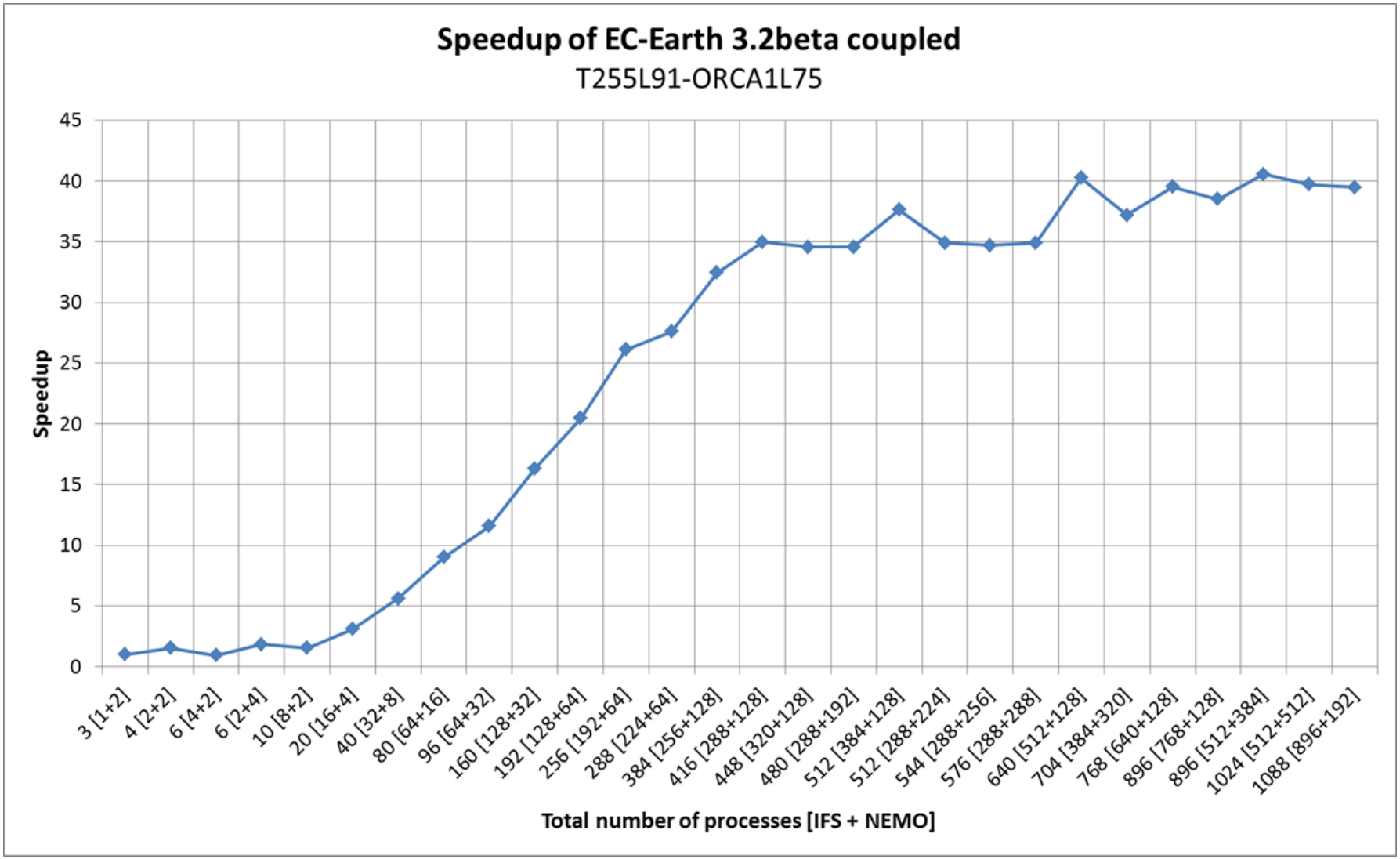
- EC-Earth 3.2
  - The Integrated Forecasting System (IFS) as atmosphere model
  - The Nucleus for European Modelling of the Ocean (NEMO) as ocean model
  - The OASIS3-MCT coupler
  - The Louvain-la-Neuve sea-Ice Model 3 (LIM3) as sea ice model



Scalability

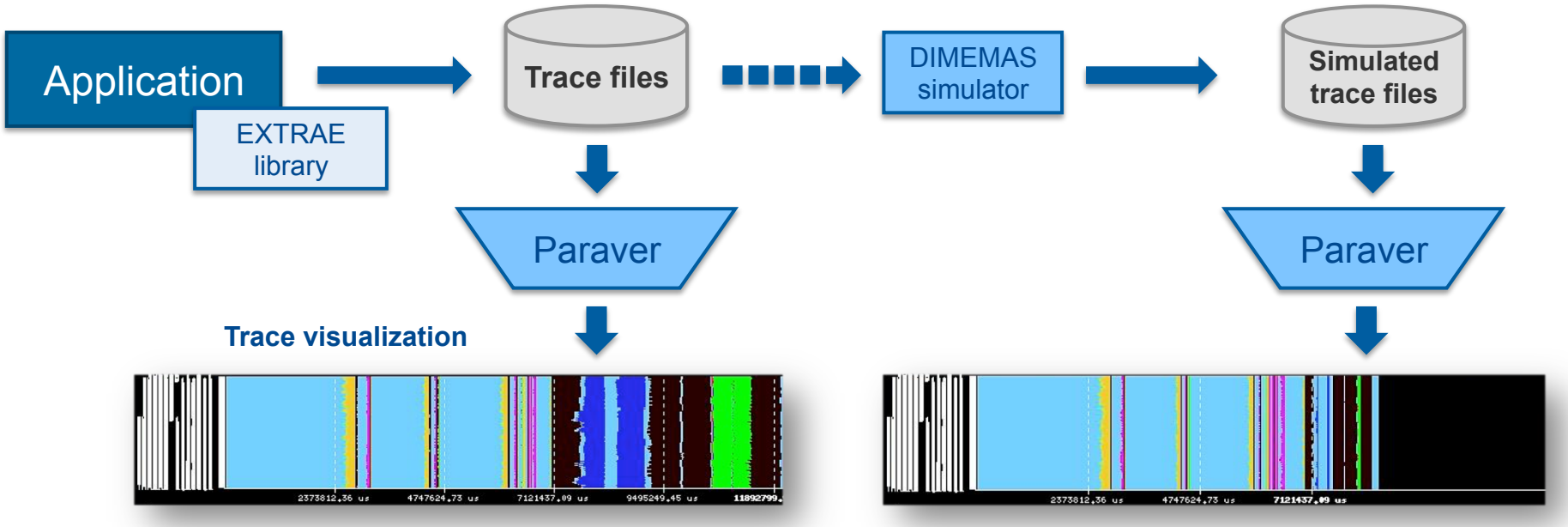


Scalability



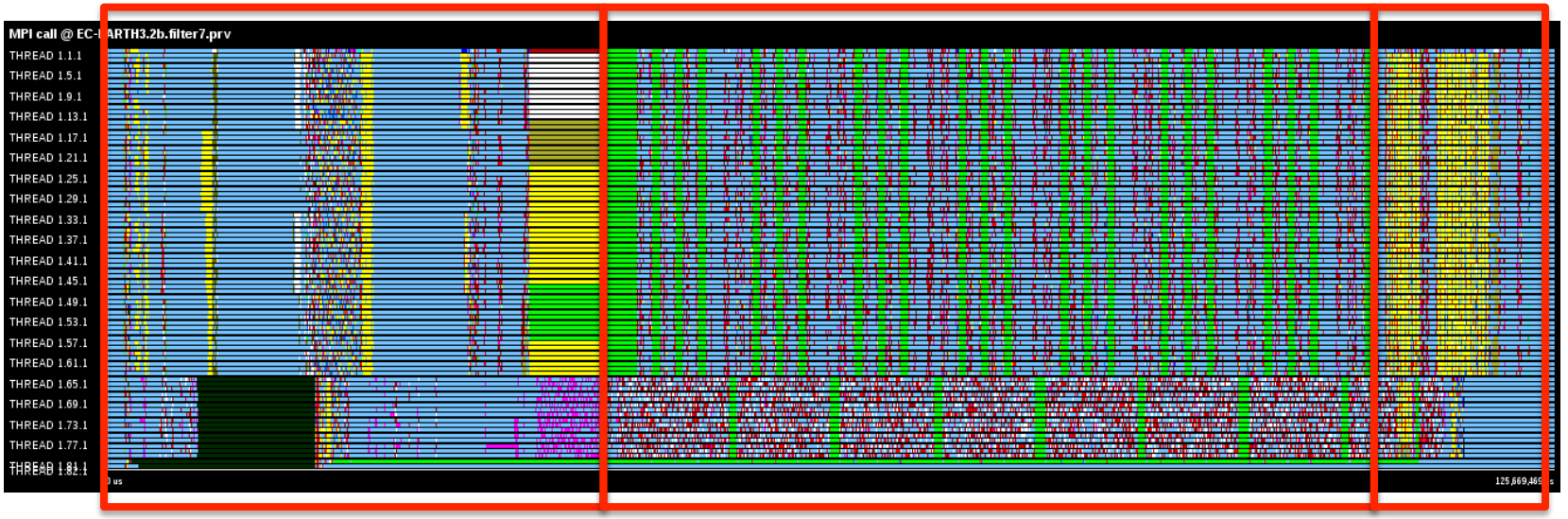
Performance tools

- Since 1991
- Based on traces
- Open Source: http://www.bsc.es/paraver
- **Extrae**: Package that generates Paraver trace-files for a post-mortem analysis
- **Paraver**: Trace visualization and analysis browser
  - Includes trace manipulation: Filter, cut traces
- **Dimemas**: Message passing simulator

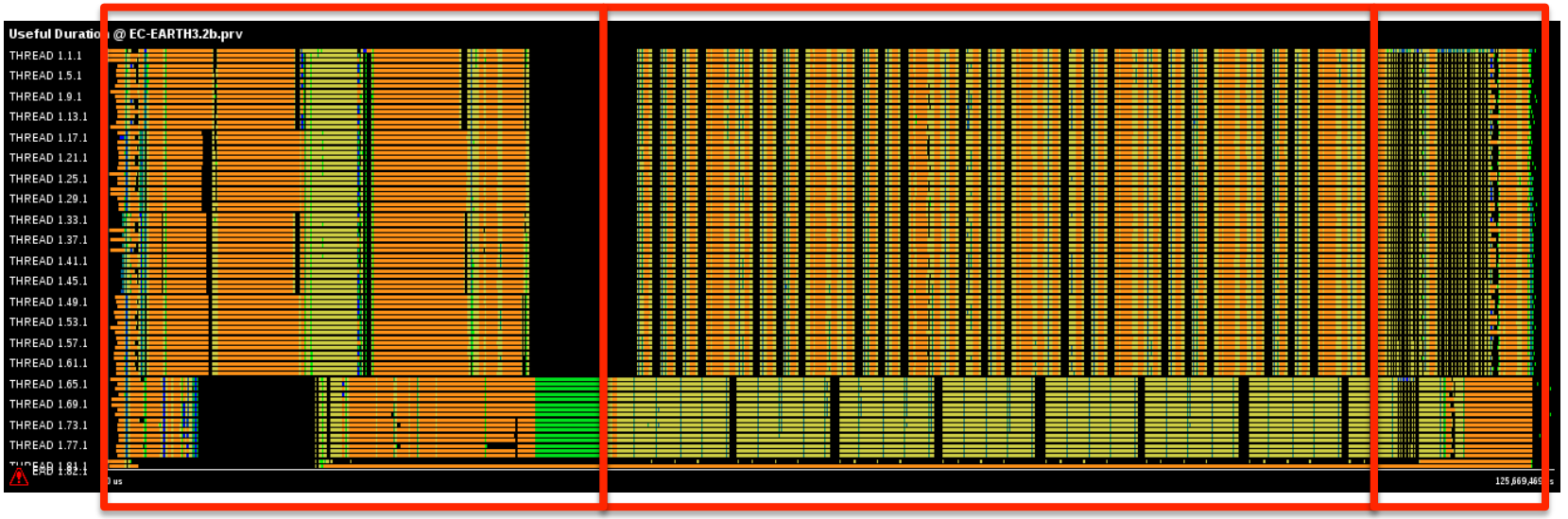


DIMEMAS generated trace. Target = ideal machine

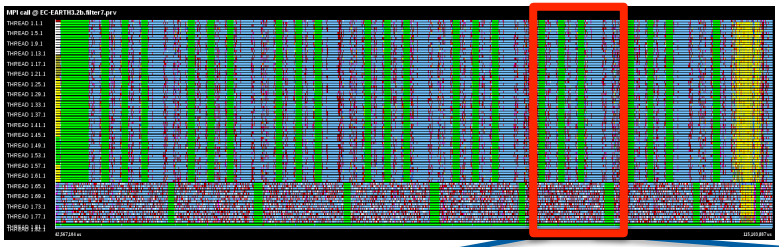
Structure of EC-Earth



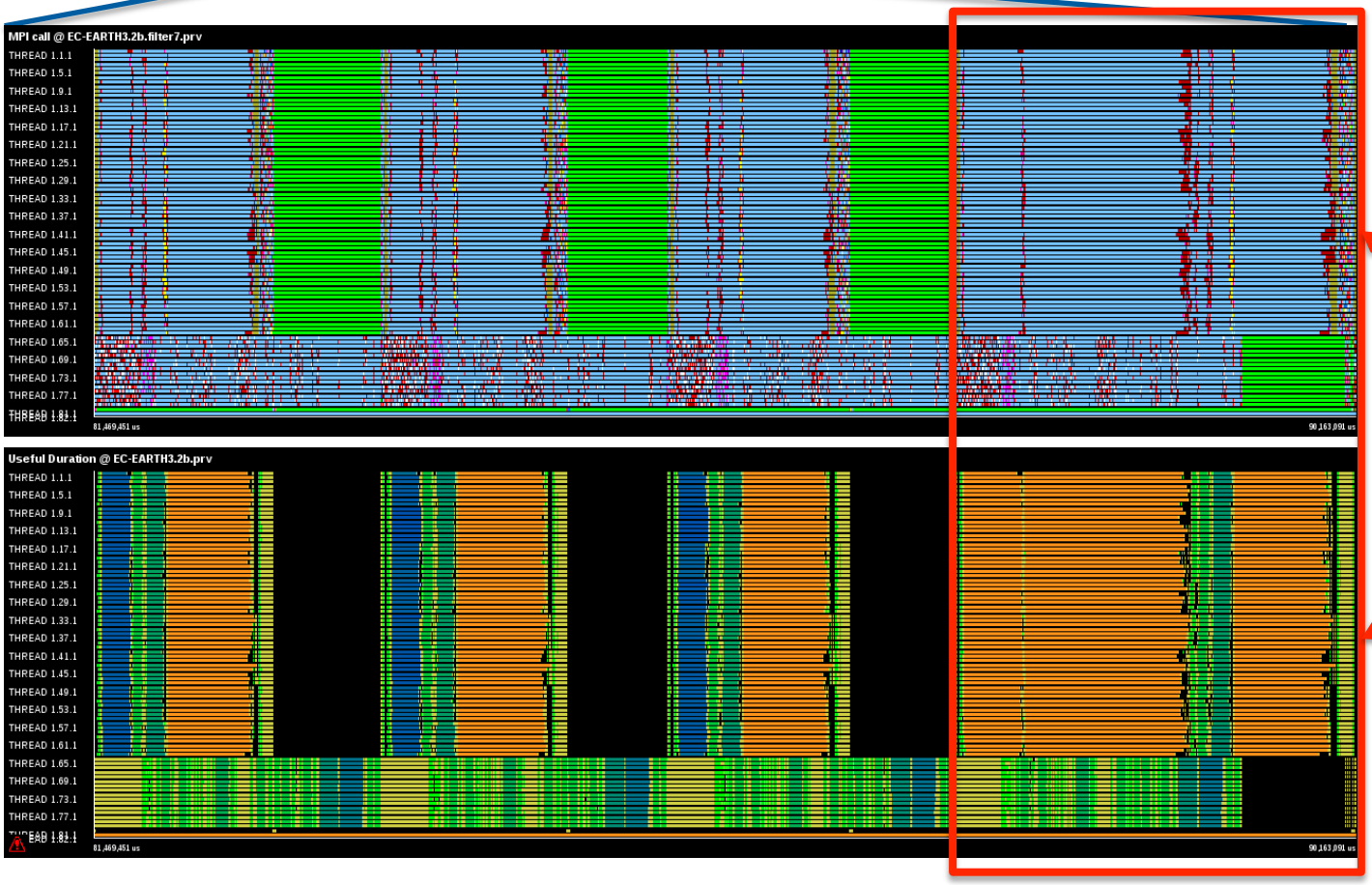
Initialization Time steps Finalization



Types of EC-Earth time steps



1 of every 4 time steps, IFS executes radiation routines, where NEMO has to wait



Time step with radiation

NEMO time step: MPI call profile

Barcelona Supercomputing Center
Centro Nacional de Supercomputación
EXCELENCIA SEVERO OCHOA

Head NEMO processes

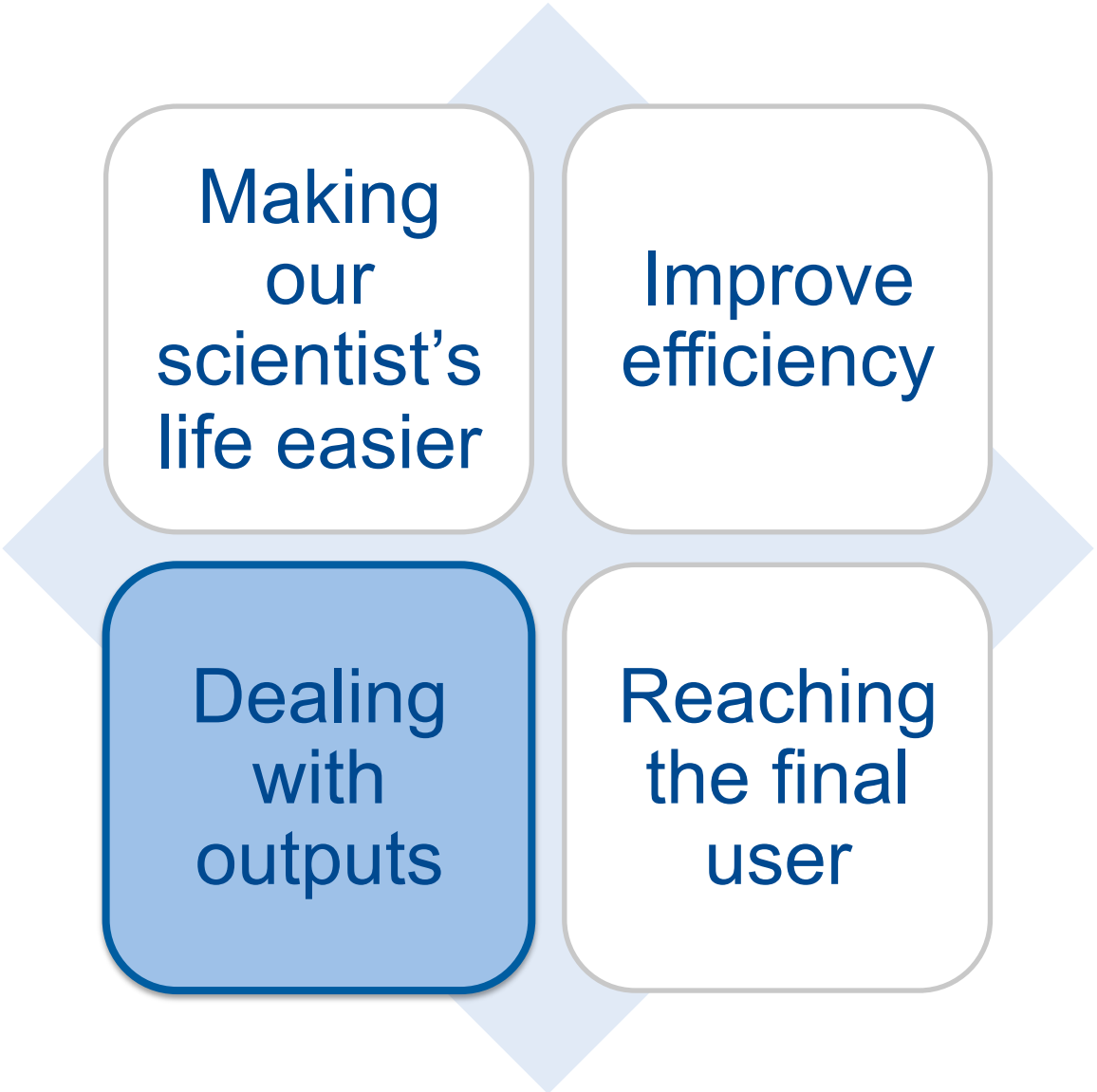
	Outside MPI	MPI_Recv	MPI_Isend	MPI_Irecv	MPI_Wait	MPI_Waitall	MPI_Allreduce
THREAD 1.513.1	42.52 %	32.87 %	7.29 %	0.05 %	3.68 %	2.28 %	10.63 %
THREAD 1.514.1	41.86 %	26.71 %	7.13 %	0.06 %	3.67 %	2.30 %	17.50 %
THREAD 1.515.1	42.48 %	27.54 %	7.31 %	0.07 %	3.75 %	2.30 %	15.82 %
THREAD 1.516.1	43.92 %	27.11 %	7.20 %	0.11 %	3.62 %	2.31 %	14.98 %
THREAD 1.517.1	47.15 %	24.67 %	7.33 %	0.11 %	3.76 %	2.31 %	13.93 %
THREAD 1.518.1	47.58 %	29.44 %	7.38 %	0.08 %	3.67 %	2.31 %	8.79 %
THREAD 1.519.1	46.25 %	29.91 %	7.34 %	0.08 %	3.76 %	2.31 %	9.60 %
THREAD 1.520.1	45.67 %	30.58 %	7.41 %	0.09 %	3.80 %	2.31 %	9.39 %
THREAD 1.521.1	45.75 %	30.90 %	7.35 %	0.09 %	3.67 %	2.31 %	9.20 %
THREAD 1.522.1	46.18 %	29.88 %	7.46 %	0.08 %	3.73 %	2.31 %	9.69 %
THREAD 1.523.1	47.53 %	30.15 %	7.27 %	0.08 %	3.71 %	2.31 %	8.24 %
THREAD 1.524.1	46.64 %	30.62 %	7.26 %	0.09 %	3.68 %	2.31 %	8.70 %
THREAD 1.525.1	44.83 %	31.21 %	7.42 %	0.08 %	3.80 %	2.31 %	9.63 %
THREAD 1.526.1	44.79 %	30.94 %	7.59 %	0.10 %	3.86 %	2.31 %	9.71 %
THREAD 1.527.1	44.14 %	31.83 %	7.35 %	0.08 %	3.66 %	2.31 %	9.89 %
THREAD 1.528.1	43.59 %	31.97 %	7.69 %	0.10 %	3.82 %	2.32 %	9.87 %
THREAD 1.529.1	45.28 %	28.35 %	9.21 %	0.14 %	4.69 %	2.28 %	9.42 %
THREAD 1.530.1	44.92 %	21.91 %	9.17 %	0.09 %	4.77 %	2.31 %	16.12 %
THREAD 1.531.1	45.82 %	22.89 %	9.01 %	0.11 %	4.65 %	2.31 %	14.53 %
THREAD 1.532.1	46.83 %	21.60 %	9.50 %	0.10 %	5.08 %	2.31 %	13.92 %

Tail NEMO processes

THREAD 1.630.1	51.72 %	23.59 %	11.11 %	0.10 %	6.18 %	2.33 %	4.28 %
THREAD 1.631.1	48.74 %	24.64 %	11.26 %	0.11 %	6.20 %	2.33 %	6.12 %
THREAD 1.632.1	47.08 %	28.74 %	9.40 %	0.11 %	4.95 %	2.33 %	6.66 %
THREAD 1.633.1	47.89 %	25.20 %	11.38 %	0.12 %	6.28 %	2.33 %	6.07 %
THREAD 1.634.1	49.82 %	24.28 %	10.98 %	0.12 %	6.21 %	2.34 %	5.57 %
THREAD 1.635.1	50.81 %	23.94 %	11.06 %	0.20 %	6.20 %	2.33 %	4.88 %
THREAD 1.636.1	50.98 %	24.12 %	10.96 %	0.09 %	6.14 %	2.33 %	4.71 %
THREAD 1.637.1	51.11 %	24.51 %	11.04 %	0.06 %	6.23 %	2.33 %	3.99 %
THREAD 1.638.1	49.27 %	25.55 %	11.08 %	0.06 %	6.44 %	2.33 %	4.57 %
THREAD 1.639.1	47.39 %	26.76 %	11.08 %	0.04 %	6.26 %	2.33 %	5.34 %
THREAD 1.640.1	47.89 %	26.38 %	11.06 %	0.00 %	6.26 %	2.33 %	5.98 %
Total	5,874.86 %	3,607.40 %	1,178.48 %	13.89 %	611.62 %	296.33 %	1,148.00 %
Average	45.90 %	28.18 %	9.21 %	0.11 %	4.78 %	2.32 %	8.97 %
Maximum	52.74 %	34.39 %	12.76 %	0.22 %	7.47 %	2.77 %	17.50 %
Minimum	41.86 %	21.60 %	7.13 %	0.00 %	3.62 %	2.27 %	3.47 %
StDev	1.92 %	2.99 %	0.98 %	0.03 %	0.66 %	0.04 %	2.91 %
Avg/Max	0.87	0.82	0.72	0.49	0.64	0.83	0.51

The main problem of NEMO are the small chunks of computation and the huge amount of communications. It has a poor efficiency of 45.9%

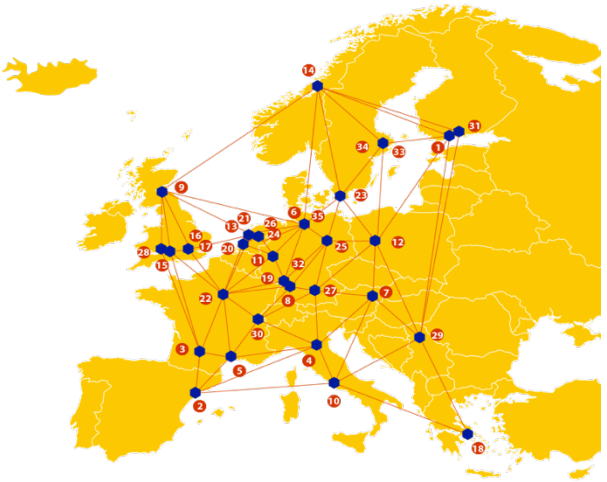
Some challenges



International Data Initiatives

- EUDAT Pilot
  - "Support to scientific research on seasonal-to-decadal climate and air quality modelling"
    - *Data sync & exchange*
    - *Data replication*
    - *Data discovery & search*
    - *Data repository & sharing*
    - *Data staging*

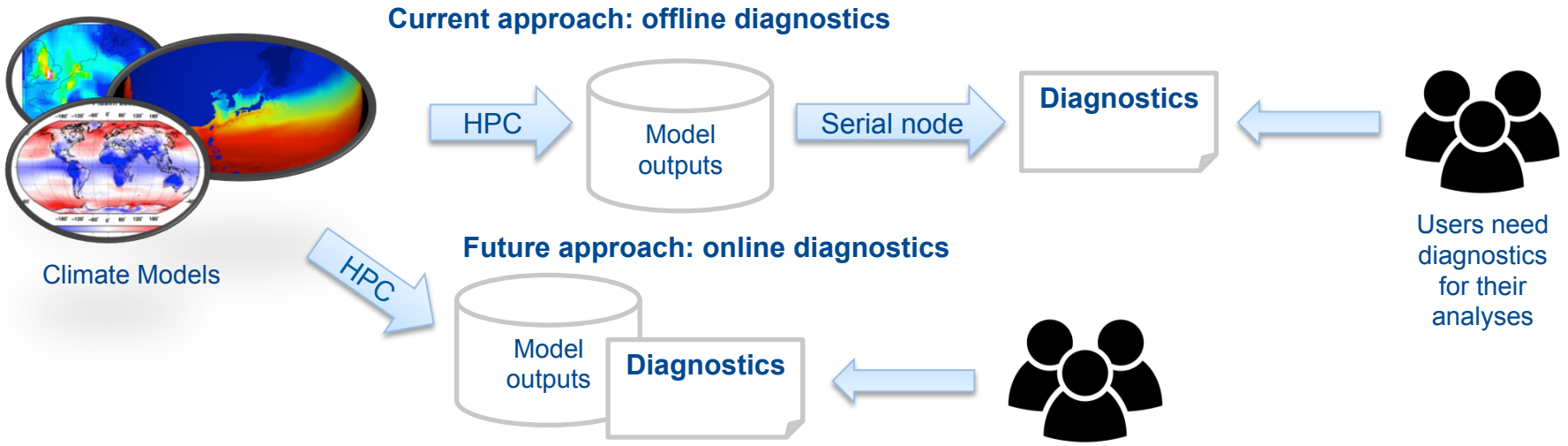
EUDAT



- RDA Interest Group in "Weather, climate and air quality"
  - Discuss the challenges for the use and efficient analysis of large and diverse datasets from the climate, weather and air quality communities
  - Strong pressure from a large user community
  - Mailing list: earthsciences-rda-ig@bsc.es

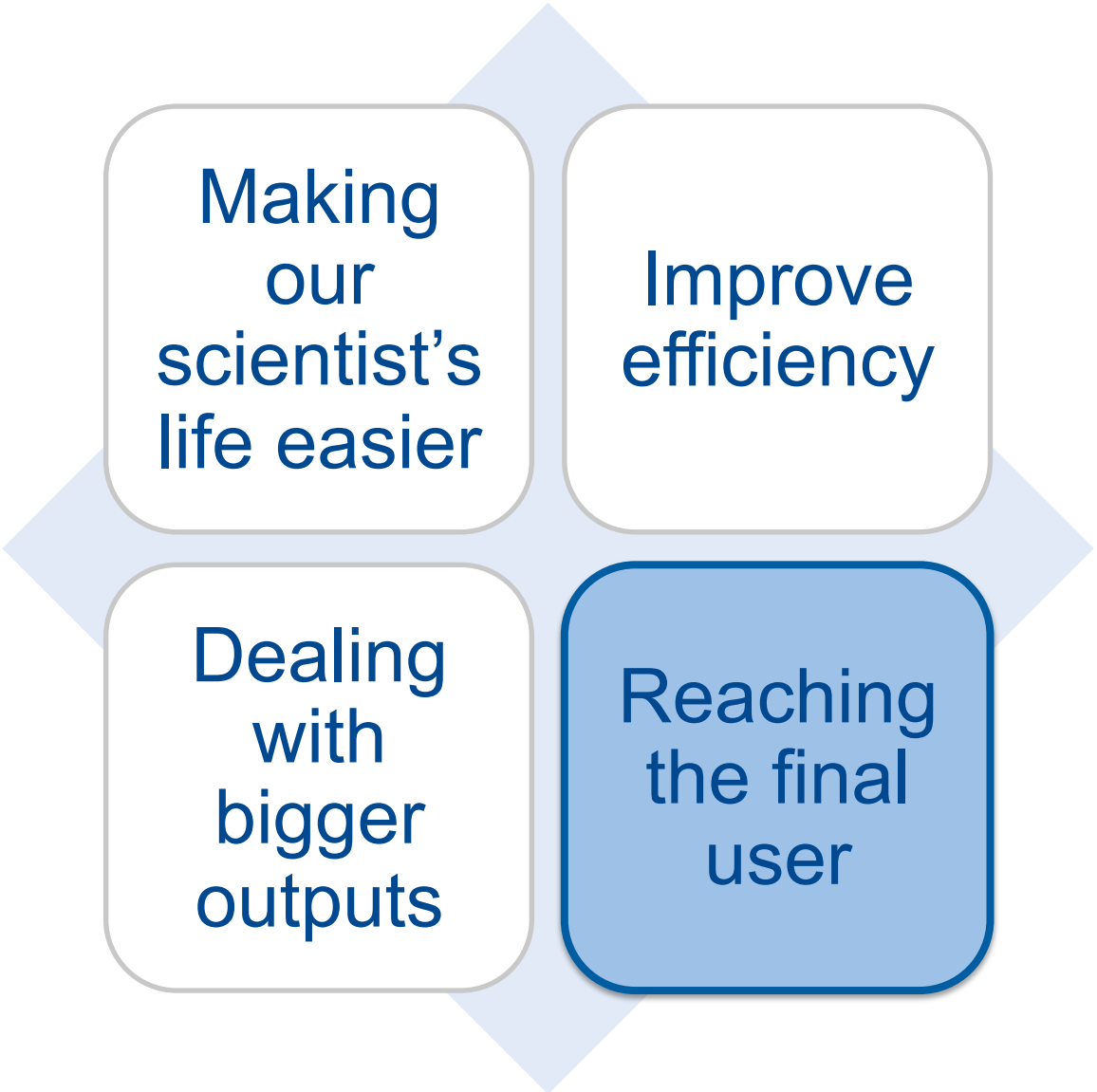
RDA
RESEARCH DATA ALLIANCE

Online diagnostics for climate models



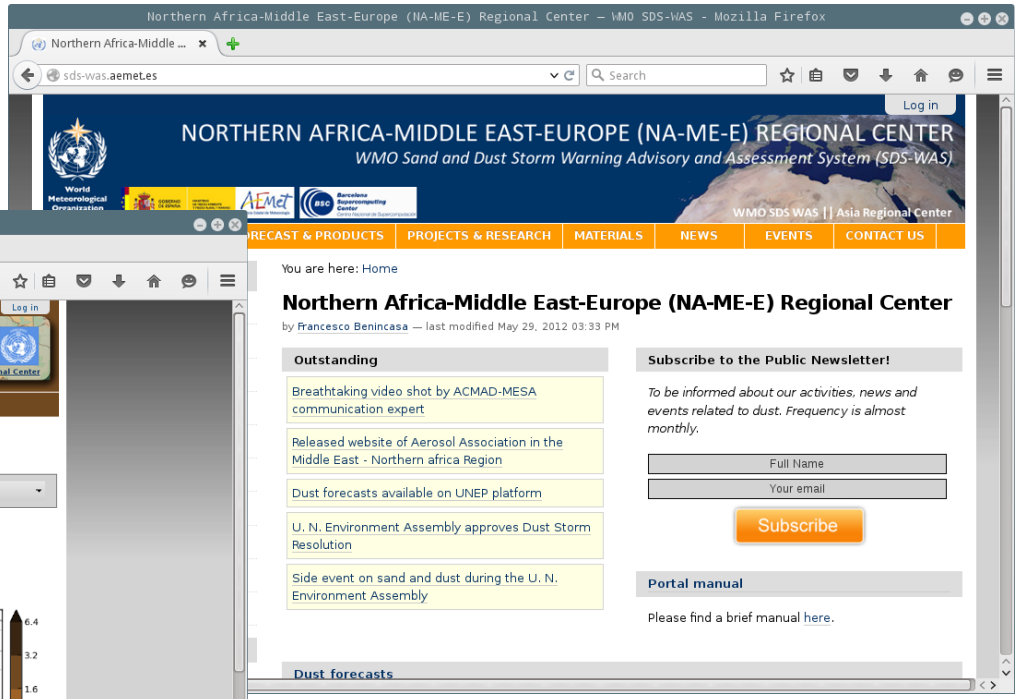
- Diagnostics computed as Analytics as a Service
  - Diagnostics online (during model run)
  - Reduced data traffic
  - Diagnostics possible on the computing nodes
  - New diagnostics (data mining of extremes) possible
  - The user gets the results faster ◊ crucial to adapt to climate change and to develop climate services (public and private)

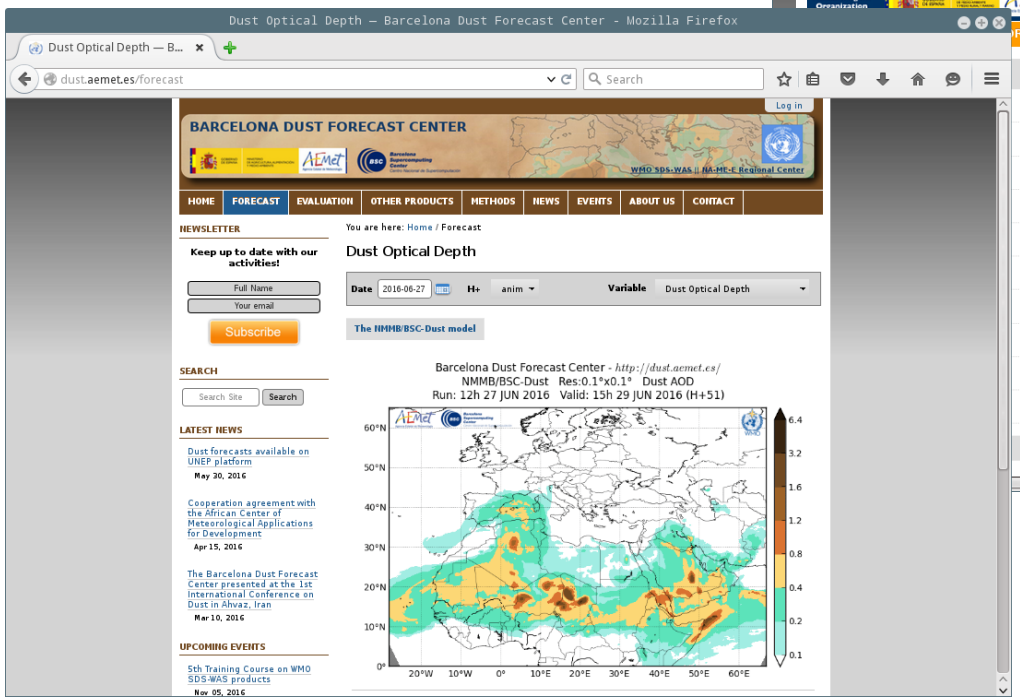
Some challenges



Operational dust forecasts

- Barcelona Dust Forecast Center (BDFC) and Sand and Dust Storm-Warning and Advisory System (SDS-WAS) for North Africa, Middle East and Europe, both operated jointly by BSC-CNS and AEMET
  - http://dust.aemet.es
  - http://sds-was.aemet.es





CALIOPE Operational Air Quality Forecast

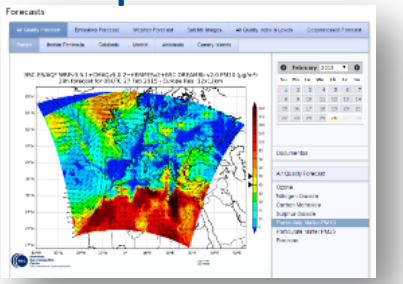
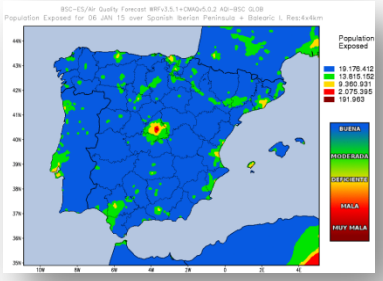
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

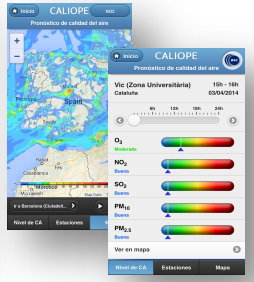
EXCELENCIA SEVERO OCHOA

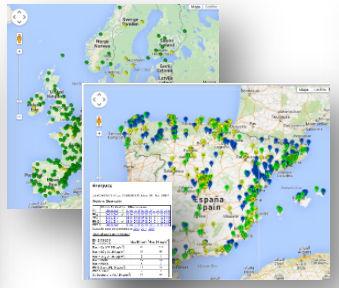
AQF CALIOPE system: daily forecast and evaluation

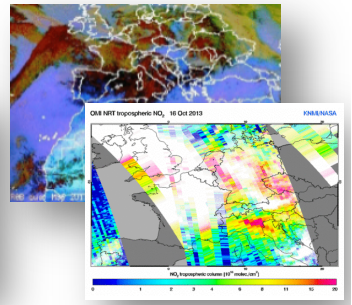
Daily forecast for **meteorology, emissions and air quality**: Europe (12km), Iberian Peninsula (4km), Andalusia, Catalonia and Madrid (1km), since 2007

Forecast products

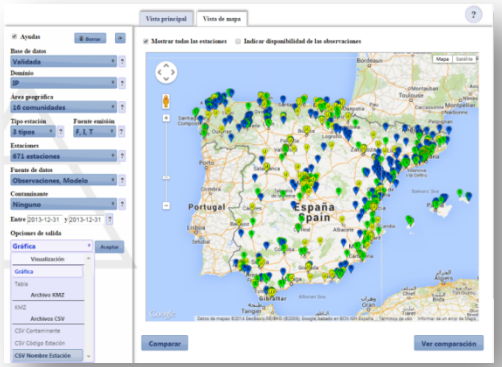





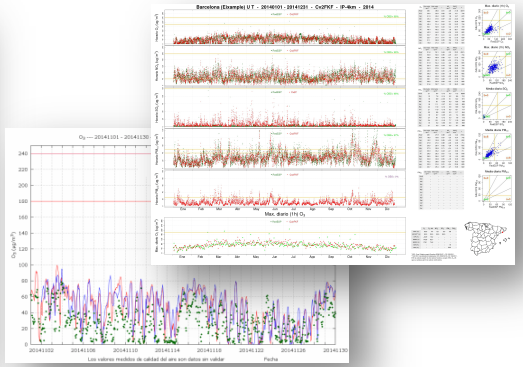




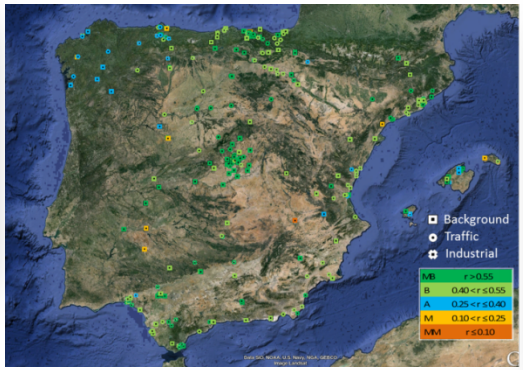
Air quality database

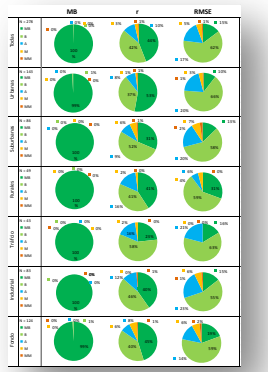


Near Real Time evaluation

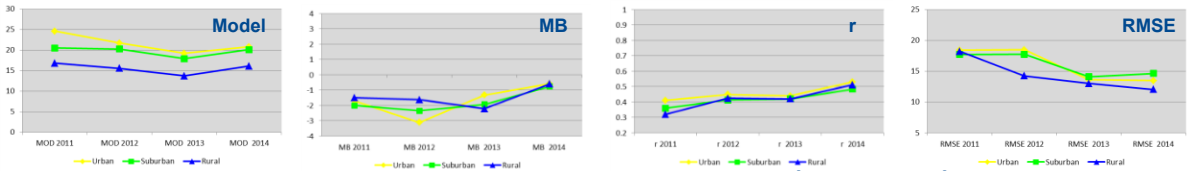


Annual evaluation by air quality stations



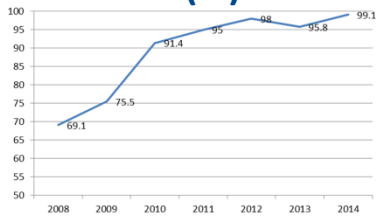


Annual follow up

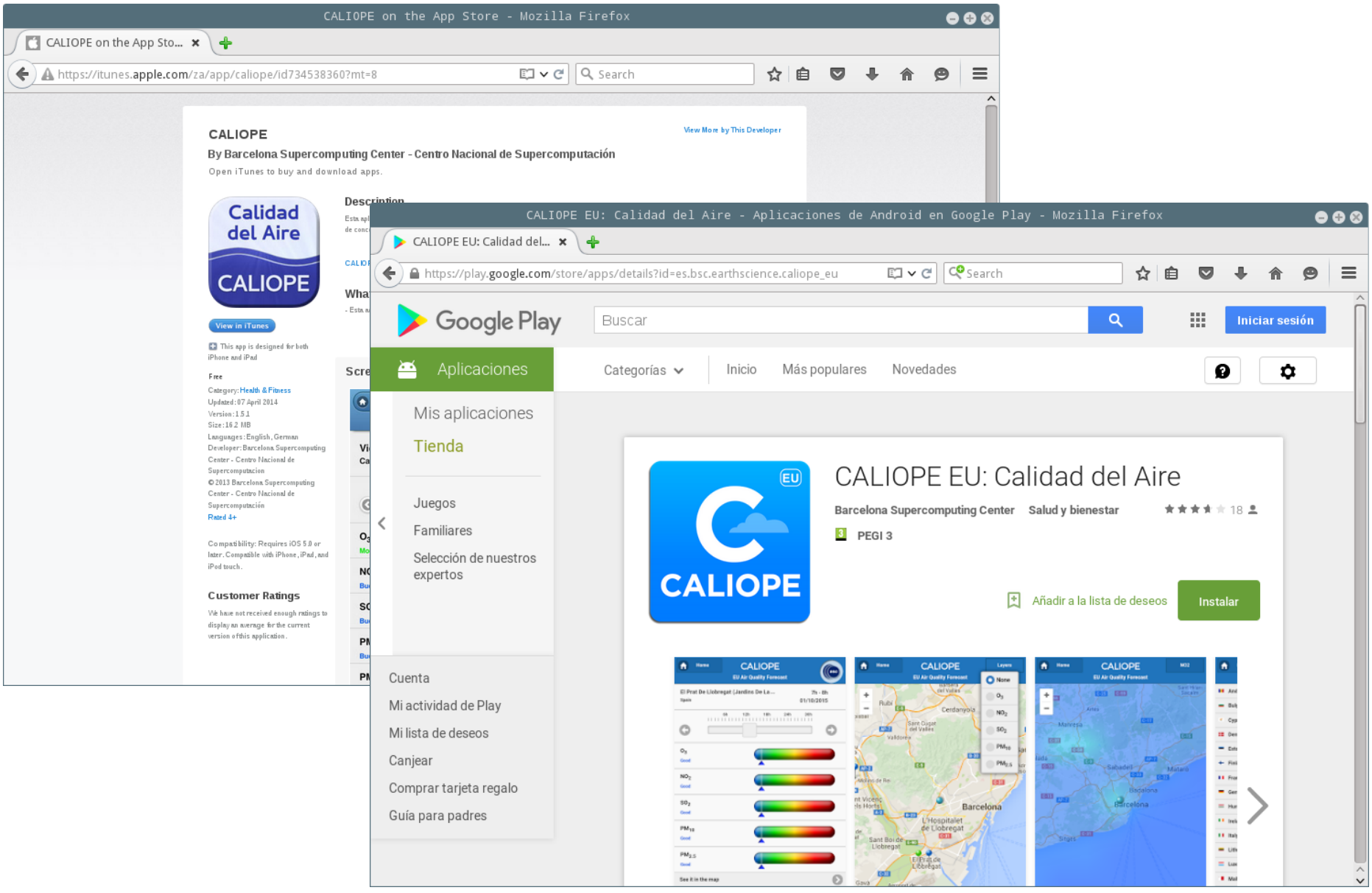


PM10_KF annual average skill evolution (2011-2014)

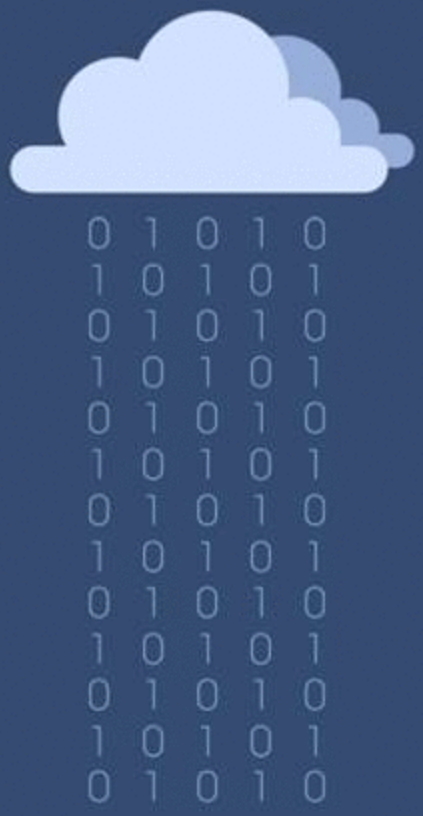
Rate of successfully completed simulations (%)



Mobile applications



Questions



www.bsc.es

Thank you!

For further information please contact
kim.serradell@bsc.es